JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 02
2025

Research Article
https://doi.org/10.56979/902/2025

Translation to a Scarce-Resource Language through Iterations

Aasim Ali[1], Sania Umer[2], Ahsan Iqbal[1], and Ashfaq Ahmad[3*]

[1]Multan University of Science and Technology, Multan, Pakistan.
[2]COMSATS University Islamabad, Wah Cantt, Pakistan.
[3]Lahore Garrison University, Lahore, Pakistan.
[*]Corresponding Author: Ashfaq Ahmad. Email: ashfaq@lgu.edu.pk

Received: June 30, 2025 **Accepted:** August 28, 2025

Abstract: Machine translation for low-resource languages such as Urdu remains a significant challenge due to limited parallel corpora and the absence of linguistic annotation tools. This study presents an iterative statistical machine translation (SMT) approach that incrementally improves translation quality using only existing bilingual text. In each iteration, the translation output of the previous model is reused as the source side to retrain a new SMT system aligned with the original target sentences. The process continues until translation quality, measured by Bilingual Evaluation Understudy (BLEU) score, stabilizes. Experiments on an English–Urdu parallel corpus demonstrate that the proposed method achieves notable improvements over the baseline system without employing any morphological or syntactic pre-processing. These findings suggest that iterative retraining can partially capture implicit linguistic patterns from limited data, offering a viable path towards improving translation for scarce-resource languages.
Keywords: Urdu; Linguistic Improvement; Scarce-resource

1. Introduction

Recent trend in machine translation is mostly towards data-driven methods including Statistical Machine Translation (SMT), which uses parallel text. This approach learns translation through phrase alignments [1] which are based on word alignments. In the seminal paper [2] of SMT it was admitted that morphological and syntactic annotations in the parallel text may improve the quality of translation. Morphological information improves learnability for realizing the correct shape of words, especially for morphologically rich languages like Arabic and Urdu. Syntactic information improves positioning of words in the given context, especially when source and target pair has different positions for grammatical relations (Subject, Object, etc.) like English versus Japanese/ Urdu. An intuitive way of algorithmic evaluation of translation output is based on the number of matching sequences and subsequences of words in comparison with human translation. We have used BLEU [3] for a quick evaluation of progress in translation improvement. A freely available toolkit for training and decoding of SMT systems is Moses [4], along with supportive tools for intermediate tasks like text alignment [5].

A limited number of language pairs have parallel text available for SMT research. The available data for English-Urdu pair is scarce [6]. Language pairs with scarce parallel text have been experimented with additional resources to improve translation, including but not limited to bilingual lexicon and morpho-syntactic annotations [7]. Manually annotating such data is extremely costly.

Statistical machine translation based approaches take the language as a collection of sentences, which are sequences of words and punctuation, divided by a tokenization step. Tokens may be tagged by their part-of-speech or by other such features that add to their grammatical associations, in terms of their shape and

arrangement. Unlike English, Urdu exhibits rich inflectional morphology [8], which results in a greater number of surface forms against a root word. The use of information related to morphology and syntax improves translation quality [7] [9] [10]. Syntactic information has also been used to build language model for target side to show improvement [11].

The contemporary models of feature factors [12] conveniently support adding linguistic information as parameters through annotation along the words in plain text. However, there is no claim on best ratio of mixture of linguistic knowledge into the word mapping model. Our experiments have shown that there was no specific upper or lower bound of progress by using several of the above mentioned methods. We have used lemmatization, part-of-speech, and stemming features to study their effect on translation results. In this paper, we propose a method of iterative improvement in the accuracies by automatic learning of these hidden aspects in the surface form of the word itself.

In the proposed method, the system gradually learns these linguistic elements (shapes and orders of words, etc.) from the surface forms of the target side, without any explicit knowledge, hint and tagging. There is no need of mono-lingual resources as addendum to the parallel text. This approach also eliminates any pre- and post-processing steps to incorporate such linguistic knowledge into the plain text.

Our output is not the linguistic information rather the learning of these elements is evident from the correctness of generated text. Those elements were neither specified in the parallel text nor are the target of our work. We have improved the shapes and arrangements of words on the target side by using the SMT process iteratively, to incrementally learn such information from simply the parallel text itself.

The first SMT model is trained for English-to-Urdu, which is used to produce translation of English. The produced translation is a crummy Urdu; rather an intermediate stage let's name it Urdu' (Urdu prime). Urdu' is no more English, but has not reached Urdu. In the next step, Urdu' is used as a source side for training another model, Urdu'-to-Urdu model produces a translation of Urdu' towards Urdu. Yet the produced translation may not have reached Urdu, and is simply the next intermediate stage; let's name this new working stage as Urdu'' (Urdu double prime), and repeat the above procedure. These steps are repeated till the output keeps improving as measured with the help of BLEU score [3] measurement on Held-out data at the end of each such iteration.

Our preliminary experiments report BLEU scores exceeding 90 on our in-house splits; however, further validation on unseen data is required. Such a high quality of correct shape and arrangement could not be achieved by using explicit morpho-syntactic annotations in our intermediate experiments. Since this approach involves no linguistic or language-specific steps, hence may be useful for any language pair.

The rest of the paper has been organized into the following sections. Section 2 gives a quick review of the existing work on usage of monolingual data and incremental learning for the sake of improving machine translation. Section 3 details the methodology of the proposed iterative algorithm. Section 4 describes the data, experimental setup, and results. Section 5 discusses the data and output patterns to advocate the proposed technique. There are interesting patterns in the translated output, which give even better results, at times, when compared manually. Discussion also records the intuition behind the new way of using the existing tools of translation, and builds a new machine that outperforms all existing ways of translation, in terms of automatic evaluation, at least.

2. Literature Review

Recent work on low-resource machine translation has investigated iterative and pseudo-labeling strategies similar in spirit to our method. Iterative and back-translation variants were revisited for low-resource pairs and shown to require careful balancing of synthetic and authentic data [60], [61]. Additionally, studies have explored iterative self-correction with large language models and token-level self-correction to improve noisy pseudo-parallel data [62], [63]. Recent surveys and empirical analyses of low-resource MT further highlight how monolingual augmentation and model priors affect outcomes, and caution about inflated automatic scores when train-test phrase overlap is high [64], [21].

Morphology and syntax are two salient factors of linguistic typology. Pair of typologically distant languages causes machine translation to be a challenging problem. The structural order of languages is categorized into three patterns: SVO, SOV, and VSO [13]. Restructuring of phrases and sentences, of participating languages, during the translation may improve the results. Reordering, insertion, and deletion of words are included in such restructuring effort. These efforts are induced due to differences in grammatical structures of source and target languages [14].

2.1. Syntactic information improves SMT output

Syntactic information of participating language pair improves the quality of translation. Syntactic information of source side improves translation [10], and syntactic restructuring of target side further improves the translation quality by filling the gaps in lexical coverage [15]. A classification technique has been used for improving the phrase selection of source side that improved the results for Arabic to English statistical machine translation [16]. Use of syntactic information in the language model of target side also improves the translation [11] [17]. Syntactic reordering using parse trees of source side (Arabic) by automatic derivation of unlexicalized reordering rules on the basis of word alignment in the preprocessing step improved the translation of unigram source language phrases into English [18]. Using lexicalized reordering, hierarchical phrase based system, and precedence reordering rules for the verb, the adjective, and the noun with preposition on an SVO source side (English) improves BLEU score when translating to each of the SOV target sides (Korean, Japanese, Hindi, Urdu, and Turkish) [19].

2.2. Morphological information improves SMT output

Languages vary on the basis of synthesis and fusion of morphology [20]. Urdu has more morphological patterns [8] as compared to English. German to English machine translation has been improved by using hierarchical lexicon to deal with the difference of morphology in both languages, in addition to sentence restructuring and support of disambiguated dictionary [7]. An efficient approach of finite state methods has been used for incorporating morphological knowledge in the source side for Persian-English translation [22]. Several methods of incorporating morphological information have been investigated to show improved translation from Czech to English [9], and from Arabic to English [23] [24]. However, mapping the morphologically simple language to a morphologically rich language is more difficult than its opposite side of translation [Lopez2008]. Morphological generation technique is used with language model of morphologically richer language on the target side, to increase the correctness of translations [25]. Improvement in translation quality and time efficiency has been reported for English to German SMT by using dependency parse of source side and on the target side by splitting the compound words in hierarchical way, adding grammatical constraints, and modifying the labels of parse tree [26].

2.3. English-to-Hindi SMT

Since Urdu is morphologically and syntactically similar to Hindi [27] therefore it may be relevant to explore the literature related to Hindi language. Rule based reordering of English syntactic structure (SVO) according to Hindi (SOV) which shows an improvement in results [28]. In addition, Ramanathan et al have reported further improvement by using the semantic relations of source side (English) to produce the case markers and affixes of target side (Hindi) [29]. Their work shows a slight decrease in this improved score when they used suffix separation to incorporate richness of Hindi morphology. The rules used in this work are similar to an Interlingua based work [30] for the same language pair. English lexicon of word senses has also been used to reduce pattern ambiguity [31] that has been hypothesized in their work to improve SMT. A probability based approach to learn local reordering of children of a node in the parse tree of source text (English) has increased the translation accuracy for Hindi (target) [32]. Ahsan et al have used the linguistic knowledge from a rule based machine English to Hindi translation system to improve BLEU score of SMT [33].

2.4. English-to-Arabic SMT

Script [34], certain morphological patterns and a large portion of vocabulary of Urdu is taken from Arabic [35] therefore it is pertinent to include related instances from literature. Morphological tokenization of Arabic text reduces sparseness in data and improves translation for English-Arabic pair [36]. English-Danish language pair has been used for reordering of English to be used in English-Arabic SMT [37]. A rule-based approach for

directly reshaping of English noun phrases, preposition phrases, and verb phrases has been used to bring it closer to the target side (Arabic) [38]. Morphological knowledge has further been exploited by investigating the segmentation scheme for breaking the tokens of Arabic being the target side text of parallel English [39] [40].

2.5. Benefits of incremental approach

The scarcity of data and the complexity of the task (of finding word sequences and shapes) induced the use of statistical method to obtain best results [41]. Statistical machine translation [14], being a machine learning approach towards translation [42], is used in the proposed work. A more detailed and updated record of statistical machine translation may be found in [14]. The proposed work considers linguistic knowledge (morphology, syntax, and word sense) to be "hidden" elements and uses the iterations of machine translation in form of expectation maximization algorithm [43] without any external knowledge, to reach better output. Our output is better in terms of correctness of shapes, sequences, and senses of words. The mapping systems may take several iterations to incrementally learn in the way the humans learn the languages [44]. The proposed work considers the translated output of source language as a pivot language [45], which is then used to improve the model to gradually reach the target language, by utilizing the power of incremental learning [46-48]. Gradual learning in several iterations reduces the impact of noise and irrelevant attributes [49] for automatically learning the word mappings to generate more correct sentences as output of translation.

The approach of incremental machine translation (IMT) uses the knowledge of human translator for enhancing the confidence of correct translations, and using that confidence for future translations [50]. The proposed work uses the same idea of enhanced confidence with the help of automatic tool (BLEU, instead of human translator) for evaluation of translated output of one pass to be used as input for translation of next pass. Daybelge and Cicekli have used a similar approach of using BLEU score as a measure of incremental learning and reported improvement in the translation quality using example based machine translation [51]. Quality of translation does not depend only on the syntax and morphology but also on the sense of the source word [52] [53].

Incremental machine translation for English-Japanese improves translation for spoken language translation, where the meaning of "incremental" is the addition of words into the input sentence on incremental basis [54]. Winiwarter has designed an incremental system of learning rules from user given bilingually parallel examples for improvement of Japanese-English translation [55]. Explicit information about syntactic and morphological structures is not readily available in text, and has to be added in the preprocessing step [14]. Using the translation of phrases observed previously increases the translation correctness when they occur subsequently [56] [57]. This is another view of "incremental" learning in which already observed high probability mappings helps improving the mappings of other translation units in subsequent passes of learning. We have successfully experimented and introduced a technique that gradually learns the linguistic information from parallel text in several iterations of translation, which is detailed in the next section.

3. Incremental technique for SMT

The argument for including syntactic annotation in statistical machine translation models rests on various points, such as, (a) Reordering for syntactic reasons, (b) Better explanation for function words, (c) Conditioning on syntactically related words, (d) Use of syntactic language models, and (e) Reduce the rate of unknown words.

The argument against syntactic annotation is that this type of markup does not occur in sentences originally, and has to be added using automatic tools. These tools can have a significant error rate so while syntax may be useful in theory, it may not be available in practice. Also, adding syntactic annotation makes models more complex (especially due to reordering) and harder to deal with (for instance, increasing search errors).

We want to directly optimize translation performance. We use the statistical machine translation methods to progress through training several stages of translations step by step. One realization of such training is incremental approach to machine translation.

The SMT learner learns the mapping probabilities between source and target words and phrases. Then SMT decoder uses those probabilities to reach the highest probable translation. This translated text has already used the good features learnt in the learning stage. These good features (correctly mapped words and phrases) help learn more good features in the next iteration of SMT learner. It improves the mapping for already-poorly-mapped words and phrases, by keeping the already-confidently-mapped words and phrases stuck to their earlier good mappings. The word mapping file produced during the training phase of Model 1 has 104,218 entries, whereas the same component of Model 3 has only 38,409 entries. It shows that the scattered mappings of 1st learning have been converged to almost 1/3rd of the mappings. An extract of some of the highest probability mappings from these files is exhibited in the Appendix A. An extract of how the immature mappings in the start happen to converge in the later stage is displayed in Appendix B.

We propose to use for extraction of implied information about morphology and syntax are the word mappings. The selection of good features (word mappings) is assumed in the parallel data. For the first iteration, the parallelization is between source side and the target side of parallel corpus. For the next step the learning is improved by the parallelization between the output of the first pass and the target. The learning keeps improving over such passes. That's why the threshold for the termination of this loop is *no further improvement*. The BLEU score is used to automatically investigate whether to carry on to next iteration or not. An important point is that mapping is learnt from Training data set while the threshold is observed on held out data set.

Those words that have found their correct mapping already in the first pass (due to high probabilities of translation and language models) will not be affected in the second pass. The second pass will only improve the probabilities of features which could not reach to their correct mapping in the target side. In this way we use the data and the translation mechanism for the selection of good features.

3.1. Learning Method

Following is the flowchart like diagram showing the outline of the proposed method (Figure 1).

The proposed learning method is an automatic process of incremental training. It is incremental in terms of learning more and more of language typology (including morpho-syntactic information) on every subsequent pass of training and translation. Figures 1.1, 1.2, and 1.3, give detailed view of abstract steps of Figure 1.

First step (1.1) in this technique is to compute the baseline model for statistical machine translation (SMT), which learns probabilities from plain bilingual parallel text. It is termed as "Model1" herein. A variable, *i*, is used to refer to the current model number therefore initialized by 1 in the start to refer to Model1. Figure 1.1 shows that original source and target sides of the parallel text are used to learn 1st model.

Second step (1.2) of Figure 1 uses Model*i*, where *i*=1 in the first iteration, and keeps incrementing in the subsequent iterations. In Figure 1.2, when *i* equals 1 (which means first pass) then original source side of held-out text (DS) is provided as input to the decoder that uses Model1 to produce the translated version of held-out data. This translation is termed as DT′ (which means 1st translation of held-out text), and is used for computing BLEU score in comparison with original target side of held-out text. DT′*(i)* is expanded as DT′ for 1st iteration, DT′′ for 2nd iteration, and so on.

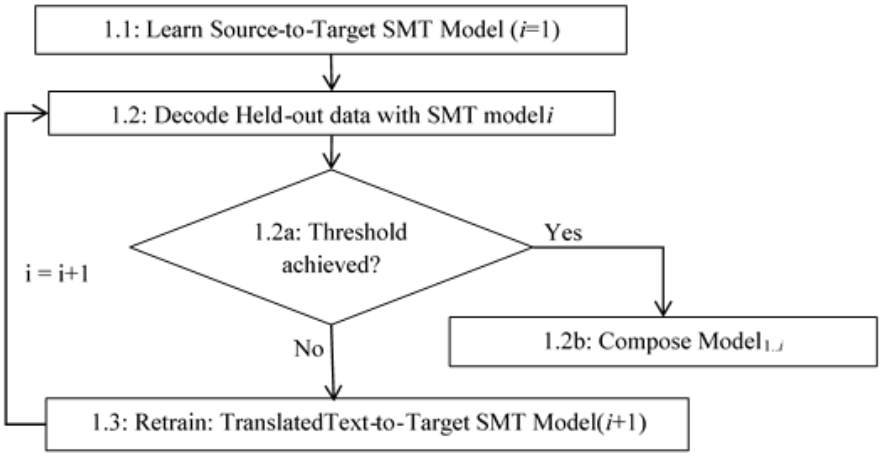

Figure 1. Outline of incremental learning procedure

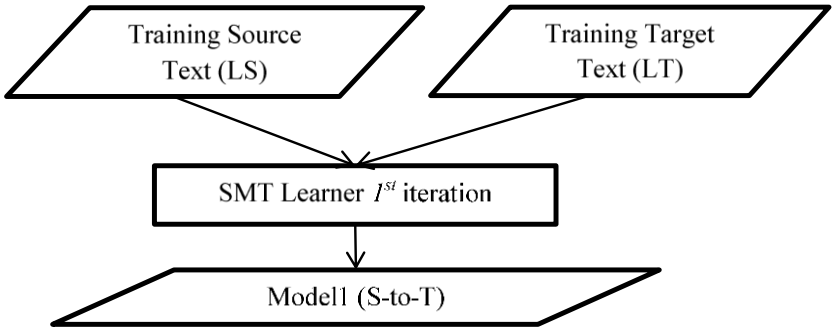

Figure 2. Learn Source-to-Target SMT Model (i=1)

In Figure 1.2, when *i* is greater than 1 (which means any non-first iteration), the translation of source side of held-out text computed in the previous iteration ((*i*-1)th translated version) is provided as input to the decoder that uses Model*i* to produce the *i* th translated version of held-out data. This translation is termed as DT'' (which means 2nd translation of held-out text) if *i*=2. It is used for computing BLEU score in comparison with original target side of the held-out parallel text.

The conditional step 1.2a of Figure 1 is simply the investigation of BLEU score for threshold condition, to decide if the next iteration will continue or not. If the threshold is achieved then all the models(1..*i*) are composed to be used for testing in the step 1.2b, as shown in experiment-specific diagram labeled as Figure 2 below (see section 4).

In the step 1.3, in case the threshold is not achieved, is to train the next model (termed as Retrain in Figure 1). In Figure 1.3, when *i* equals 1 (which means first pass) then original source side of training text (LS) is provided as input to the decoder that uses Model1 to produce the translated version of training data. Translated text is LT' (which means 1st translation of training text) in 1st iteration which is used as source side for Model2 (as denoted by Model*i*+1 in Figure 1.3). Now, the target side of plain parallel text is still the same as original, however the source side is the *i* th translation of source text. LT'*(i)* is expanded as LT' for 1st iteration, LT'' for 2nd iteration, and so on.

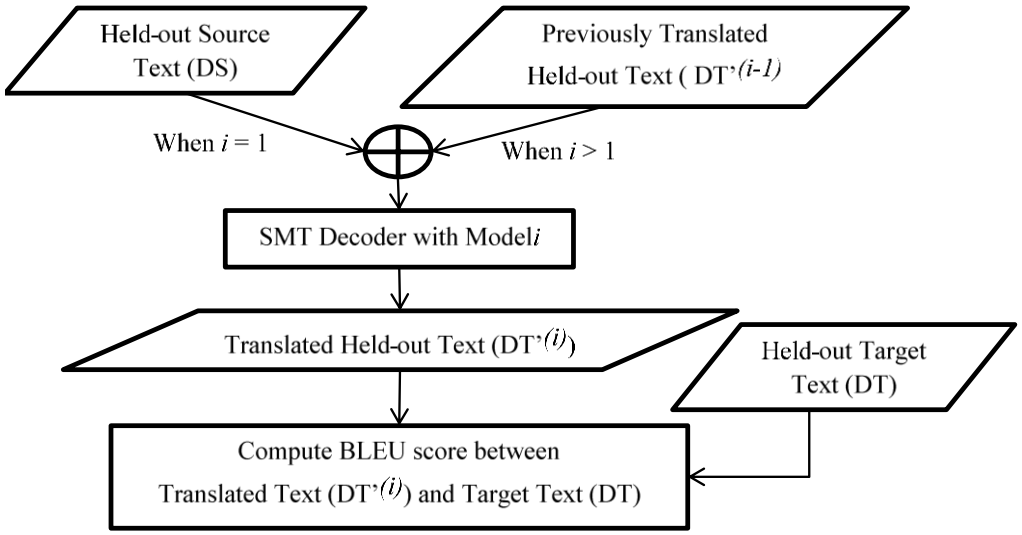

Figure 3. Decode held-out data with SMT Modeli, to compute BLEU score for threshold

In Figure 1.3, when *i* exceeds 1 (which means any non-first iteration), the translation of source side of training text computed in the previous iteration ((*i*-1)th translated version) is provided as input to the decoder that uses Model*i* to produce the *i* th translated version of training data. This *i* th translation is used as input to learn Model*i*+1.

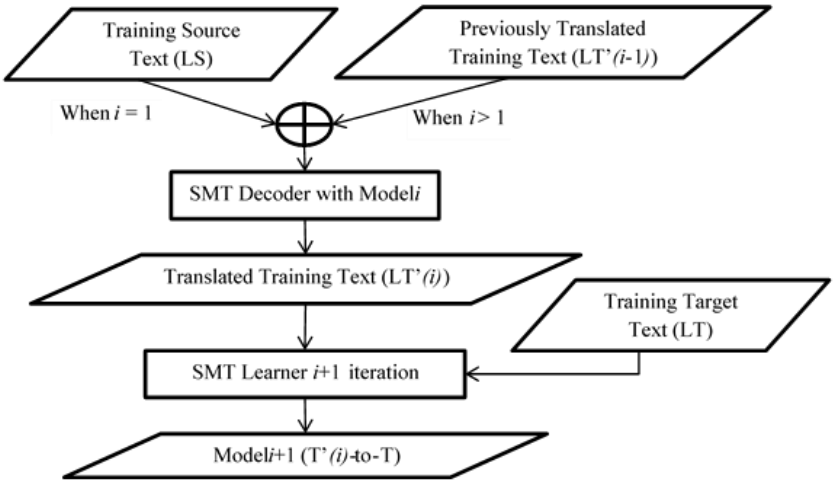

Figure 4. Retrain TranslatedText-to-Target SMT Modeli+1, for next iteration

For example, if *i*=2, the LT' (means 1st translation of training text, computed in previous iteration) is an input to the decoder which produces LT'' (means 2nd translation of training text, denoted as LT'(*i*) in Figure 1.3) which is used as source side for learning Model3 (as denoted by Model*i*+1 in Figure 1.3). Now, the target side of plain parallel text is still the same as original, however the source side is the *i* th translation of source text.

4. Verifying experiments

This section details the experiments to verify the success of proposed method. The first subsection (4.1) describes the data, second subsection (4.2) shows the results of baseline experiment, third subsection (4.3) notes the results of using morpho-syntactic annotations, concluding subsection (4.4) describes the results and examples illustrating the improvement of output. Bilingual parallel corpus is the fundamental resource for SMT. we have used English-Urdu parallel text of [58], which is sentence level aligned.

4.1. Data

Text from two books is used in this study. The English and Urdu versions of these books are already aligned at topic level (containing one or more paragraphs). There are 497,354 words in 26,822 sentences on English side and 513,550 words in parallel Urdu translations. Issues related to availability of parallel data, alignment of sentences, adjustment of punctuations and bullets, and the translation differences due to morphological richness (e.g. honor) have been described in the literature [8].

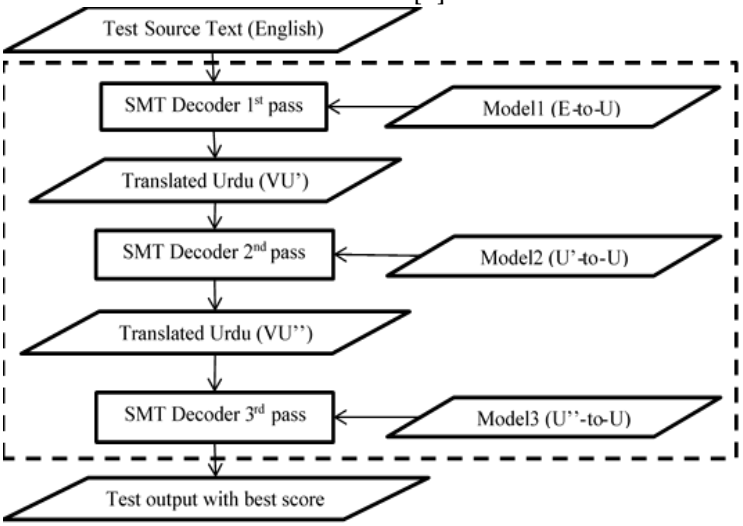

Figure 5. Detailed view of applying the iteratively learnt and tuned model on Test data. Dashed box outlines the model which is composed of Model1, Model2, and Model3. This model has learnt shapes and sequences of words from the running text of source side input, which is very close to the reference Urdu.

4.2. Experiment and result

This experiment is designed to test if the un-annotated text can itself incrementally take the desired shapes and sequences induced by the implicit morpho-syntactic knowledge which is always present in the running text. Following is the abstraction of this new algorithm or procedure:

1. Executed the training model of baseline, i.e. English-Urdu (E-U) Model, on the training set itself (to prepare an intermediate train set let's say LU'), and obtained the BLEU score of 48.7, which is the highest score so far. The translation of held-out data from this E-U Model is also saved and termed as DU' to be used in the next stage.
2. The translated training source (LU') was paired with the original Urdu target (U). A new model (U'-U) was trained to learn morpho-syntactic information not captured in the first-pass E-U model. When run on the baseline-translated held-out set (DU'), an improvement in the BLEU score was observed. This process yielded the next version of the held-out translations, DU''.
3. Executed the training model U'-U on the LU' itself (to prepare another intermediate train set let's say LU'') for next stage of learning.
4. We used the latest translated training portion (LU'') with the original Urdu target (U) to train a further model, U''-U. This step aims to capture morpho-syntactic information that remained unresolved after the second pass. When evaluated on the translated held-out set (DU''), the model improved the BLEU score. It achieved a highest BLEU of 95.19.
5. As a further proof, executed the U''-U training model on step-wise prepared separately kept test data (let's call it VU''), to obtain the verification of the highest BLEU score, which gave 92.41 score.

The above obtained scores are not improved any further due to certain unchanged probabilities, e.g. اللہ تعالیٰ vs. اللہ. Some examples of incremental improvements are given in the next section. Summary of results shown in Table 1 below clearly shows that incremental learning proposed in this paper gives the highest BLEU score.

Table 1. Summary of BLEU score for English-Urdu SMT using various techniques

Experiment	BLEU
Decoding EvalSet with Training Model of Baseline TrainSet Corpora	32.10
Decoding EvalSet with Tuned Model of Baseline DevSet Corpora	37.10
Decoding EvalSet with Training Model of Morpho-Syntactic Annotated TrainSet Corpora	36.73
Decoding DevSet with Training Model of *Incremental Learning* on Baseline TrainSet Corpora	95.19
Decoding eValSet with Training Model of *Incremental Learning* on Baseline TrainSet Corpora	92.41

Table 2 demonstrates two examples that illustrate improvement of translation due to iterative process of incremental learning. Analysis of source-target mapping lexicon verifies the insights. There are 104,218 entries in the English-Urdu model (Model1 in the Figure 2) and only 38,409 entries in Urdu''-Urdu model (Model3 in Figure 2). Some entries are shown in the appendices to illustrate the convergence of word mapping due to the incremental approach.

5. Discussion and Conclusion

Table 1 given in previous section shows that both combinations of different data for testing produce the score above 90. No other combination could reach even 40 score, due to insufficient size of the parallel corpus to learn that level of morpho-syntactic information for correct mappings. This level of score has never been achieved for any other language pair even by using millions of parallel sentences. One reason of unprecedentedly high score under proposed technique is the significant overlap of phrases in the train and test data sets. However, it is also important to keep in mind that gain from this overlap could not be exploited without using the power of incremental learning [46] [47] [48] [49].

There are many such phrases in the system generated translation table that have multiple possible targets against one source. The context may not be a clue for such different choices made by human, e.g., "Allah" might have been translated as "اللہ" or as "اللہ تعالٰی". Since the underlying model probabilistically selects only one of multiple such choices, and the evaluation mechanism matches the words (not their meaning); therefore overall accuracies remained below 100.

Table 2. Examples from data, demonstrating the improvement in word positions and translations.

Following example demonstrates the improvement in *positions (reordering)* of words (reference sentence# 32):	
English Source	And Asma' bint Abu Bakr cut a piece of her girdle and tied the mouth of the leather bag with it. That is why she was called Dhat-an-Nitaqaln.
Human Translation	اسماء بنت ابی بکر نے اپنا ڈوپٹہ پھاڑا اور اس سے تھیلی کا منہ باندھ دیا، اسی وجہ سے ان کو ذات النطاق کہا جاتا ہے۔
Output from Incremental Model	اسماء بنت ابی بکر نے اپنا ڈوپٹہ پھاڑا اور اس سے تھیلی کا منہ باندھ ان کو ذات النطاق کہا جاتا ہے۔
Intermediate Output	اسماء بنت ابی بکر نے اپنا ڈوپٹہ پھاڑا اور تھیلی کا منہ باندھ اس سے ان کو ذات النطاق کہا جاتا ہے۔
Following example demonstrates the improvement in ***translation and shapes*** of words (reference sentence# 4):	
English Source	Then he washed his forearms and passed his wet hands over his head.
Human Translation	پھر اپنے دونوں ہاتھ دھوئے، سر کا مسح کیا۔
Output from Incremental Model	پھر اپنے دونوں ہاتھ دھوئے ، سر کا مسح کیا۔
Intermediate Output	پھر آپ نے اپنا forearms اور پر مسح کیا۔ سر
Following example demonstrates the improvement in ***selection of translation*** (reference sentence # 472):	
English Source	I went, and behold! It was Abu Bakr.
Human Translation	میں گیا تو دیکھا کہ حضرت ابوبکر تھے
Output from Incremental Model	میں گیا تو دیکھا کہ حضرت ابوبکر تھے
Intermediate Output	میں گیا اور دیکھا کہ حضرت ابوبکر تھے

Since this approach involves no language-specific steps therefore it may be applied to any language pair. The technique of exploiting the overlapping in TrainSet and EvalSet by iteratively learning of morphology and syntax, without human annotation, may work well for translation of any other text that typically has significant overlap of phrases including user manuals, blogs, specific news genre, and research articles from a specific field. It may also be applied for word sense disambiguation (WSD) using parallel corpus, instead of using explicit linguistic knowledge to resolve the word sense [59].

References

1. Koehn, P., Och, F. J., & Marcu, D. (2003, May). Statistical phrase-based translation. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1 (pp. 48-54). Association for Computational Linguistics.
2. Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., ... & Roossin, P. S. (1990). A statistical approach to machine translation. Computational linguistics, 16(2), 79-85.
3. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 311-318). Association for Computational Linguistics.
4. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., ... & Dyer, C. (2007, June). Moses: Open source toolkit for statistical machine translation. In Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions (pp. 177-180). Association for Computational Linguistics.
5. Och, F. J., & Ney, H. (2003). A systematic comparison of various statistical alignment models. Computational linguistics, 29(1), 19-51.
6. Ali, Aasim, Shahid Siddiq, and M. Kamran Malik. (2010). Development of parallel corpus and English to urdu statistical machine translation. International Journal of Engineering and Technology/IJENS, 01:30–33. ISSN 2077-1185.
7. Nießen, S., & Ney, H. (2004). Statistical machine translation with scarce resources using morpho-syntactic information. Computational linguistics, 30(2), 181-204.
8. Ali, Aasim. (2010). Study of Morphology of Urdu Language, for Its Computational Modeling: Study of Morphological Patterns in Urdu Language, and Partial Implementation of Computational Solution for the Same Using a Finite State Tool. VDM Publishing. ISBN 9783639289442.
9. Goldwater, S., & McClosky, D. (2005, October). Improving statistical MT through morphological analysis. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 676-683). Association for Computational Linguistics.
10. Yamada, K. and Knight, K. (2001). A syntax-based statistical translation model. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics (ACL '01). Association for Computational Linguistics, Stroudsburg, PA, USA, 523-530. DOI=10.3115/1073012.1073079 http://dx.doi.org/10.3115/1073012.1073079
11. Charniak, E., Knight, K., & Yamada, K. (2003, September). Syntax-based language models for statistical machine translation. In Proceedings of MT Summit IX (pp. 40-46).
12. Koehn, P., and Hoang, H. (2007). Factored Translation Models, In Proceedings of EMNLP-2007.
13. Greenberg, J. (1966) "Some Universals of Grammar with Particular Reference to The Order Meaningngful Eelements," in J. Greenberg, Universals of Language. MIT Press, Cambridge, Mass.
14. Koehn, P. (2010). Statistical Machine Translation (1st ed.). Cambridge University Press, New York, NY, USA.
15. Ambati, V., Lavie, A., & Carbonell, J. (2009). Extraction of syntactic translation models from parallel data using syntax from source and target languages. Proceedings of the 12th Machine Translation Summit, 190-197.
16. España-Bonet, C., Giménez, J., & Màrquez, L. (2009). Discriminative phrase-based models for Arabic machine translation. ACM Transactions on Asian Language Information Processing (TALIP), 8(4), 15.
17. Marcu, D., Wang, W., Echihabi, A., & Knight, K. (2006, July). SPMT: Statistical machine translation with syntactified target language phrases. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 44-52). Association for Computational Linguistics.
18. Habash, N. (2007). Syntactic preprocessing for statistical machine translation. MT Summit XI, 215-222.
19. Xu, P., Kang, J., Ringgaard, M., & Och, F. (2009, May). Using a dependency parser to improve SMT for subject-object-verb languages. In Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics (pp. 245-253). Association for Computational Linguistics.
20. Pirkola, A. (2001). Morphological typology of languages for IR. Journal of Documentation, 57(3), 330-348.
21. Hsu, B., Currey, A., Niu, X., Nădejde, M., & Dinu, G. (2023). Pseudo-Label Training and Model Inertia in Neural Machine Translation. arXiv preprint arXiv:2305.11808.
22. Amtrup, J. W. (2003). Morphology in machine translation systems: Efficient integration of finite state transducers and feature structure descriptions. Machine Translation, 18(3), 217-238.

23. El Isbihani, A., Khadivi, S., Bender, O., & Ney, H. (2006, June). Morpho-syntactic Arabic preprocessing for Arabic-to-English statistical machine translation. In Proceedings of the Workshop on Statistical Machine Translation (pp. 15-22). Association for Computational Linguistics.
24. Habash, N., & Sadat, F. (2006). Arabic preprocessing schemes for statistical machine translation.
25. Minkov, E., Toutanova, K., & Suzuki, H. (2007, June). Generating complex morphology for machine translation. In ACL (Vol. 7, pp. 128-135).
26. Williams, P., Sennrich, R., Nadejde, M., Huck, M., Hasler, E., & Koehn, P. (2014, June). Edinburgh's Syntax-Based Systems at WMT 2014. In Proc. of the Workshop on Statistical Machine Translation (WMT), Baltimore, MD, USA (pp. 207-214).
27. Mukund, S., Srihari, R., & Peterson, E. (2010). An Information-Extraction System for Urdu---A Resource-Poor Language. ACM Transactions on Asian Language Information Processing (TALIP), 9(4), 15.
28. Ramanathan, A., Hegde, J., Shah, R. M., Bhattacharyya, P., & Sasikumar, M. (2008, January). Simple Syntactic and Morphological Processing Can Help English-Hindi Statistical Machine Translation. In IJCNLP (pp. 513-520).
29. Ramanathan, A., Choudhary, H., Ghosh, A., & Bhattacharyya, P. (2009, August). Case markers and morphology: addressing the crux of the fluency problem in English-Hindi SMT. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2 (pp. 800-808). Association for Computational Linguistics.
30. Dave, S., Parikh, J., & Bhattacharyya, P. (2001). Interlingua-based English–Hindi Machine Translation and Language Divergence. Machine Translation, 16(4), 251-304.
31. Chatterjee, N., Goyal, S., & Naithani, A. (2005, September). Resolving pattern ambiguity for english to hindi machine translation using WordNet. In Workshop on Modern Approaches in Translation Technologies.
32. Visweswariah, K., Navratil, J., Sorensen, J., Chenthamarakshan, V., & Kambhatla, N. (2010, August). Syntax based reordering with automatically derived rules for improved statistical machine translation. In Proceedings of the 23rd international conference on computational linguistics (pp. 1119-1127). Association for Computational Linguistics.
33. Ahsan, A., Kolachina, P., Kolachina, S., Sharma, D. M., & Sangal, R. (2010). Coupling statistical machine translation with rule-based transfer and generation. In Proceedings of the 9th Conference of the Association for Machine Translation in the Americas.
34. Mirdehghan, M. (2010). Persian, Urdu, and Pashto: A comparative orthographic analysis. Writing Systems Research, 2(1), 9-23.
35. Khan, A. G., & ALWARD, M. A. (2011). ARABIC LOAN WORDS IN URDU: A LINGUISTIC ANALYSIS. Speech Context, 13-20.
36. Zollmann, A., Venugopal, A., & Vogel, S. (2006, June). Bridging the inflection morphology gap for Arabic statistical machine translation. In Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers (pp. 201-204). Association for Computational Linguistics.
37. Elming, J., & Habash, N. (2009, March). Syntactic reordering for English-Arabic phrase-based machine translation. In Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages (pp. 69-77). Association for Computational Linguistics.
38. Badr, I., Zbib, R., & Glass, J. (2009, March). Syntactic phrase reordering for English-to-Arabic statistical machine translation. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (pp. 86-93). Association for Computational Linguistics.
39. El Kholy, A., & Habash, N. (2012). Orthographic and morphological processing for English–Arabic statistical machine translation. Machine Translation, 26(1-2), 25-45.
40. Al-Haj, H., & Lavie, A. (2012). The impact of Arabic morphological segmentation on broad-coverage English-to-Arabic statistical machine translation. Machine translation, 26(1-2), 3-24.
41. Casacuberta, F., & Vidal, E. (2007). Learning finite-state models for machine translation. Machine Learning, 66(1), 69-91.
42. Cardie, C., & Mooney, R. J. (1999). Guest editors' introduction: Machine learning and natural language. Machine Learning, 34(1), 5-9.

43. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 1-38.
44. Suppes, P., Böttner, M., & Liang, L. (1995). Comprehension grammars generated from machine learning of natural languages. Machine Learning, 19(2), 133-152.
45. Wu, H., & Wang, H. (2007). Pivot language approach for phrase-based statistical machine translation. Machine Translation, 21(3), 165-181.
46. Kirby, S., & Hurford, J. (1997). The evolution of incremental learning: language, development and critical periods. Edinburgh Occasional Papers in Linguistics, 97(2), 1-33.
47. Giraud-Carrier, C. (2000). A note on the utility of incremental learning. AI Communications, 13(4), 215-223.
48. Solomonoff, R. J. (2002, December). Progress in incremental machine learning. In NIPS Workshop on Universal Learning Algorithms and Optimal Search, Whistler, BC.
49. Aha, D. W. (1992). Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. International Journal of Man-Machine Studies, 36(2), 267-287.
50. Toselli, A. H., Vidal, E., & Casacuberta, F. (2011). Incremental and Adaptive Learning for Interactive Machine Translation. In Multimodal Interactive Pattern Recognition and Applications (pp. 169-177). Springer London.
51. Daybelge, T., & Cicekli, I. (2011). A ranking method for example based machine translation results by learning from user feedback. Applied Intelligence, 35(2), 296-321.
52. Lee, H. A. (2006). Translation selection through machine learning with language resources. In Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead (pp. 370-377). Springer Berlin Heidelberg.
53. Carpuat, M., & Wu, D. (2007, June). Improving Statistical Machine Translation Using Word Sense Disambiguation. In EMNLP-CoNLL (Vol. 7, pp. 61-72).
54. Matsubara, S., & Inagaki, Y. (1997). Utilizing extra-grammatical phenomena in incremental English-Japanese machine translation. In Proc. of (Vol. 7, pp. 31-38).
55. Winiwarter, W. (2005). Incremental learning of transfer rules for customized machine translation. In Applications of Declarative Programming and Knowledge Management (pp. 47-64). Springer Berlin Heidelberg.
56. Bannard, C., & Callison-Burch, C. (2005, June). Paraphrasing with bilingual parallel corpora. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 597-604). Association for Computational Linguistics.
57. Callison-Burch, C., Koehn, P., & Osborne, M. (2006, June). Improved statistical machine translation using paraphrases. In Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (pp. 17-24). Association for Computational Linguistics.
58. Ali, A., Hussain, A., & Malik, M. K. (2013). Model for english-urdu statistical machine translation. World Applied Sciences, 24, 1362-1367.
59. Specia, L., Srinivasan, A., Joshi, S., Ramakrishnan, G., & Nunes, M. D. G. V. (2009). An investigation into feature construction to assist word sense disambiguation. Machine Learning, 76(1), 109-136.
60. Ahmed, M. A., Kashyap, K., Talukdar, K., & Boruah, P. A. (2023). Iterative back translation revisited: An experimental investigation for low-resource English Assamese neural machine translation. In Proceedings of the 20th International Conference on Natural Language Processing (ICON) (pp. 172-179).
61. Frontull, S., & Moser, G. (2024). Rule-based, neural and LLM back-translation: Comparative insights from a variant of Ladin. arXiv preprint arXiv:2407.08819.
62. Chen, P., Guo, Z., Haddow, B., & Heafield, K. (2023). Iterative translation refinement with large language models. arXiv preprint arXiv:2306.03856.
63. Lu, J., & Zhang, J. (2024, May). Improving Unsupervised Neural Machine Translation via Training Data Self-Correction. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 8942-8954).
64. Sälevä, J., & Lignos, C. (2024, August). Language Model Priors and Data Augmentation Strategies for Low-resource Machine Translation: A Case Study Using Finnish to Northern Sámi. In Findings of the Association for Computational Linguistics ACL 2024 (pp. 12949-12956).